THURSDAY, DECEMBER 13, 1894.

DILETTANTISM AND STATISTICS.

The Growth of St. Louis Children. By William Townsend Porter. Vol. vi. No 12: Transactions of the Academy of Science of St. Louis.

THE anthropometrical researches of Mr. Francis Galton have had a very widespread influence, and, under his inspiration, a large mass of material has been, and is being collected, which cannot fail to be of service in elucidating a number of knotty problems. In particular, very comprehensive measurements of children have been made in America by Bowditch, Sargant, and Porter, the results being tabulated by aid of Mr. Galton's method of percentiles.

It is not only in the form of statistics of measurements on man, but also of measurements on the lower animals and on plants, that material for the study of variation and correlation is accumulating with almost alarming rapidity. We say with almost alarming rapidity, for not only is theory lagging somewhat behind the needs of statistical experience, but, what is far worse, many collectors have no real insight into the theory which does exist. They are tabulating results in a form which will be of no permanent service, or neglecting to publish the very details which alone would enable us to test any development of statistical theory. The ignorance of the elements of mathematics, to say nothing of the theory of chance, is, indeed, singularly characteristic of many investigators, who think statistics can be handled, and conclusions drawn from them, without the least preliminary training. For example, in a recent paper on variation in the flowers of buttercups and clover, by Herr de Vries, we are introduced to our old friend the normal curve of errors as a definite symmetrical polygon; no attempt is made to fit it by aid of the probable error of the observations, but a particular form of it, apparently selected at random, is plumped down on the top of another polygon representing the observations. The odds against the thus selected theoretical polygon (curiously called the "Galton Curve") expressing the real distribution of variation, are, it is needless to say, many thousands to one. Still more curiously, sets of observations giving theoretically very good frequency curves of the type which occurs in infantile mortality, the valuation of houses, &c., are termed "Half-Galton-Curves," although the type, so far from being represented by the half of the normal curve of errors, corresponds to a curve asymptotic to the ordinate of maximum frequency! We might pass these eccentricities by, were not statistics thus handled used to support some vague theory of heredity, or to question some principle of variation.

Mr. Porter is not quite so wild in theory as Herr de Vries, but for him also the normal curve of errors appears to be a polygon, and he tells us, on p. 277, after a table of its ordinates—which are quoted from Thoma, and given only to the units, and *not to decimals*—"that there is a limit beyond which no deviation occurs." It is, perhaps, needless to say that if we start only with a knowledge of the theory of chance, such as may be found in works like those of Thoma and Stieda, we are hardly likely to avoid bad theoretical blunders. Mr. Porter is no worse than many writers of memoirs in the *Journal of the Royal Statistical Society;* but the time has come when statistics, as well as archæology, must be taken out of the hands of the dilettanti, and when a man, if he takes upon himself to publish statistical researches, must, like the physicist, show his credentials in the form of a fair mathematical training.

Unfortunately, Mr. Galton's method of percentiles, useful as it is in the right place, is now acting as a distinct check to statistical theory in a somewhat unexpected manner. Before that theory was disseminated the dilettante statistician was compelled to publish his raw material, and his remarks on it did not impair its value to the trained scientist. Now, however, the percentile method of dealing with statistics seems so intelligible, that the non-mathematical anthropologist or physiologist grasps it at once, reserves his raw material, and publishes tables of percentiles. It is hard to conceive anything more disastrous for true statistical progress.

The reason for this is easily explained. As is well known, any normal curve of errors can be reduced to any other by uniform stretches (or squeezes) parallel to its base and its axis. The area of any portion of any normal curve up to a deviation of given magnitude cannot be given by a finite series; it can by stretching be reduced to an integral, the values of which are tabulated, but which is not itself finitely expressible for an indefinite value of the deviation. The tables formed of this integral, however, enable us to pass from selected percentiles to the constants, and so to the form of the frequency curve. Now, innumerable frequency curves in biological, just as in economical statistics, are *not* of the normal type. A generalised frequency curve can be theoretically obtained fitting closely the observations in many, perhaps in all such cases, but owing to its asymmetry or skewness, its limited range and other features, it neither possesses the stretch property of the normal curve, nor can its areas be expressed by tables of one entry. The tables required are, according to the type, of two or three entries, and such tables are not likely to be prepared for a long time to come, and if prepared, would be of little service except for the case when the statistics are given in percentiles. In other words, the use of percentiles may suffice to demonstrate the skewness of the statistics, but at once precludes the use of any theory higher than that of the normal curve, by which the skewness could be accounted for. This is the fundamental defect of Mr. Porter's labours. He gives us a most extensive system of measurement on boys and girls from six to twenty years of age—ample material, if properly dealt with, to provide solutions for innumerable problems in correlation and selection with age. The material, however, is only given in the form of percentiles or in diagrams of the "ogive" curve corresponding to the integral of the frequency curve. These diagrams fully confirm the results of Bowditch—*that the statistics are skew, and that the degree of skewness varies with the age.* But all means of scientifically treating this skewness disappear, because we have only percentile results. The skewness of the original frequency curve cannot be deduced from nine or ten points on a curve, the equation of which is only expressible by an unintegratable form containing the con-

© 1894 Nature Publishing Group

stants to be determined. This is all the more tantalising, as the variation with age (or with selection) of the skewness of the frequency curve for a given organ is very probably a most important factor in the mathematical treatment of evolution, while it is precisely *human* material which enables us to record the age, too often an unknown quantity in investigations on the lower animals. It must not be supposed, however, that Mr. Porter realises that skewness in the percentiles marks want of normality in the frequency curve. On the contrary, he takes the raw material for the height of 2192 girls aged eight, and fits it, not very accurately, to a normal frequency distribution. Although the measurements better fit a skew than a normal distribution, the divergence from the normal is not very marked, and Mr. Porter writes:

> "It is not necessary to make this comparison at more than one age, or in more than one dimension, for it is known that if one series in a group like that with which we have to deal shows this agreement, the other series will be found to do the same." (p. 290.)

Now the whole point of the measurements, if properly interpreted, is to ascertain the varying degree of skewness with age, *i.e.* the varying amount of divergence from the normal curve. To emphasise this divergence when remarking on his raw material, but to fail to recognise what bearing it has on theory, is typical of the untrained statistician.

As we have seen, the use of percentiles renders Mr. Porter's material of small value for the problem of selection. The same remark again applies in the case of correlation. The ten organs measured would have been most valuable material for the problem of correlation; the percentile results alone being given, nothing can be done. In Chapter ix. Mr. Porter does develop a theory of correlation, which, however, seems to us absolutely erroneous. As he makes it the basis of a scheme for the school physician to testify whether the pupil's strength is equal to the strain put upon it, this want of accurate theoretical knowledge appears doubly unfortunate.

We have ventured to point out these failings in Mr. Porter's work, not because they are peculiar to him—they are characteristic of much work of a statistical kind which is now being turned out in both Europe and America—but rather because they point to a new need, which the public has hardly yet recognised. There has been up to the present time—with the honourable exception of courses of lectures by Dr. Venn in Cambridge—no teaching of statistical theory in England. There is no chair of statistics in any English university, and the Newmarch lectureship at University College, except for the year of its tenancy by Prof. Edgeworth, has been held by economists, and not by mathematically trained statisticians. We want a centre, which shall not only contain a statistical museum, but embrace as well a statistical laboratory and workshop. In such a centre students might not only receive a mathematical training in dealing with raw statistics, but also be exercised in the methods of collecting and tabulating, which must precede mathematical reduction. To such a centre the biologist and anthropometrist could send their measurements to be "fitted," the physicist his observations to be dealt with, and the economist or sociologist his price or labour statistics to be analysed. At the same time, absolute measurements might be made bearing on the problems of evolution, disease and national economy. In this manner a number of efficient young statisticians might be trained, who would not only find a life-work ready for them in craniology, zoology, botany, and economics, but who, passing into government departments, census offices and labour bureaus, might remove from us the reproach of a recent continental writer, that nowhere were statistical dilettanti so rampant as in England.

KARL PEARSON.

WATER SUPPLY AND WATER-WORKS.

The Principles of Water-works Engineering. By J. H. Tudsbery Turner and A. W. Brightmore. (London: Spon, 1893.)

The Water Supply of Towns. By W. K. Burton. (London: Crosby Lockwood and Son, 1894.)

WITH the constant growth of population, and the increasing tendency of the population to congregate into large towns, the provision of an ample supply of pure water to towns becomes every year of greater importance; whilst, as the nearer sources of supply become inadequate for the ever-growing demands, a sufficient supply becomes more difficult to obtain, and has to be sought at considerable distances. Accordingly, the subject of water-supply has assumed an enhanced importance within recent years; and till these books appeared, there was a dearth of comprehensive text-books on this branch of engineering. As the sources of water-supply, and the works for its provision and distribution are comprised within very definite limits, both books necessarily traverse much the same ground, and deal with similar classes of works; but, nevertheless, they exhibit differences in arrangement, and in the method of treating the various subjects. Thus, whereas "Water-works Engineering" extends over 420 pages, separated into only eight chapters, and illustrated by one hundred and twenty figures in the text, "The Water Supply of Towns" is subdivided into twenty-two chapters, occupying only 272 pages, somewhat larger in size, but illustrated by two hundred and fifty-seven drawings, several of which are put into forty-three large folding-plates. The first book, moreover, is subdivided into numbered sections, the numbers of which are merely used for reference, and instead of the pages in the index; whilst the various subjects in the second book are clearly indicated by black-letter headings. Altogether, "The Water Supply of Towns" is much better arranged for reference; and though containing slightly less matter, it is much more fully illustrated, is in larger print, and has a somewhat longer index than "Water-works Engineering." Neither book gives a summary of contents at the head of each chapter, which is often serviceable for reference.

In "Water-works Engineering," the sources, measurement, collection, storage, purification, conveyance, and distribution of water, and the maintenance of water-works, are successively dealt with in the eight chapters. The most logical sequence would be to describe in regular order the various processes to which water is subjected, from its source up to its delivery to the consumer.

© 1894 Nature Publishing Group